

[SEQUENCE LISTING]

SEQUENCE ID No.: 1

LENGTH: 251

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: peptide

SEQUENCE:

Met Ser Met Pro Asn Ile Val Pro Pro Ala Glu Val Arg Thr Glu Gly  
5 10 15  
Leu Ser Leu Glu Glu Tyr Asp Glu Glu Gln Val Arg Leu Met Glu Glu  
20 25 30  
Arg Cys Ile Leu Val Asn Pro Asp Asp Val Ala Tyr Gly Glu Ala Ser  
35 40 45  
Lys Lys Thr Cys His Leu Met Ser Asp Ile Asn Ala Pro Lys Asp Leu  
50 55 60  
Leu His Arg Ala Phe Ser Val Phe Leu Phe Arg Pro Ser Asp Gly Ala  
65 70 75 80  
Leu Leu Leu Gln Arg Arg Ala Asp Glu Lys Ile Thr Phe Pro Gly Met  
85 90 95  
Trp Thr Asn Thr Cys Cys Ser His Pro Leu Ser Ile Lys Gly Glu Val  
100 105 110  
Glu Glu Glu Asn Gln Ile Gly Val Arg Arg Ala Ala Ser Arg Lys Leu  
115 120 125  
Glu His Glu Leu Gly Val Pro Thr Ser Ser Thr Pro Pro Asp Ser Phe  
130 135 140  
Thr Tyr Leu Thr Arg Ile His Tyr Leu Ala Pro Ser Asp Gly Leu Trp  
145 150 155 160  
Gly Glu His Glu Ile Asp Tyr Ile Leu Phe Ser Thr Thr Pro Thr Glu  
165 170 175  
His Thr Gly Asn Pro Asn Glu Val Ser Asp Thr Arg Tyr Val Thr Lys  
180 185 190  
Pro Glu Leu Gln Ala Met Phe Glu Asp Glu Ser Asn Ser Phe Thr Pro  
195 200 205  
Trp Phe Lys Leu Ile Ala Arg Asp Phe Leu Phe Gly Trp Trp Asp Gln  
210 215 220  
Leu Leu Ala Arg Arg Asn Glu Lys Gly Glu Val Asp Ala Lys Ser Leu  
225 230 235 240

Glu Asp Leu Ser Asp Asn Lys Val Trp Lys Met \*\*\*

245

250

SEQUENCE ID No.: 2

LENGTH: 259

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: peptide

SEQUENCE:

Met Gln Leu Leu Ala Glu Asp Arg Thr Asp His Met Arg Gly Ala Ser  
5 10 15  
Thr Trp Ala Gly Gly Gln Ser Gln Asp Glu Leu Met Leu Lys Asp Glu  
20 25 30  
Cys Ile Leu Val Asp Ala Asp Asn Ile Thr Gly His Val Ser Lys  
35 40 45  
Leu Glu Cys His Lys Phe Leu Pro His Gln Pro Ala Gly Leu Leu His  
50 55 60  
Arg Ala Phe Ser Val Phe Leu Phe Asp Asp Gln Gly Arg Leu Leu Leu  
65 70 75 80  
Gln Gln Arg Ala Arg Ser Lys Ile Thr Phe Pro Ser Val Trp Thr Asn  
85 90 95  
Thr Cys Cys Ser His Pro Leu His Gly Gln Thr Pro Asp Glu Val Asp  
100 105 110  
Gln Leu Ser Gln Val Ala Asp Gly Thr Val Pro Gly Ala Lys Ala Ala  
115 120 125  
Ala Ile Arg Lys Leu Glu His Glu Leu Gly Ile Pro Ala His Gln Leu  
130 135 140  
Pro Ala Ser Ala Phe Arg Phe Leu Thr Arg Leu His Tyr Cys Ala Ala  
145 150 155 160  
Asp Val Gln Pro Ala Ala Thr Gln Ser Ala Leu Trp Gly Glu His Glu  
165 170 175  
Met Asp Tyr Ile Leu Phe Ile Arg Ala Asn Val Thr Leu Ala Pro Asn  
180 185 190  
Pro Asp Glu Val Asp Glu Val Arg Tyr Val Thr Gln Glu Glu Leu Arg  
195 200 205

Gln Met Met Gln Pro Asp Asn Gly Leu Gln Trp Ser Pro Trp Phe Arg  
210 215 220  
Ile Ile Ala Ala Arg Phe Leu Glu Arg Trp Trp Ala Asp Leu Asp Ala  
225 230 235 240  
Ala Leu Asn Thr Asp Lys His Glu Asp Trp Gly Thr Val His His Ile  
245 250 255  
Asn Glu Ala \*\*\*

SEQUENCE ID No.: 3

LENGTH: 288

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: peptide

SEQUENCE:

Met Thr Ala Asp Asn Asn Ser Met Pro His Gly Ala Val Ser Ser Tyr  
5 10 15  
Ala Lys Leu Val Gln Asn Gln Thr Pro Glu Asp Ile Leu Glu Glu Phe  
20 25 30  
Pro Glu Ile Ile Pro Leu Gln Gln Arg Pro Asn Thr Arg Ser Ser Glu  
35 40 45  
Thr Ser Asn Asp Glu Ser Gly Glu Thr Cys Phe Ser Gly His Asp Glu  
50 55 60  
Glu Gln Ile Lys Leu Met Asn Glu Asn Cys Ile Val Leu Asp Trp Asp  
65 70 75 80  
Asp Asn Ala Ile Gly Ala Gly Thr Lys Lys Val Cys His Leu Met Glu  
85 90 95  
Asn Ile Glu Lys Gly Leu Leu His Arg Ala Phe Ser Val Phe Ile Phe  
100 105 110  
Asn Glu Gln Gly Glu Leu Leu Gln Gln Arg Ala Thr Glu Lys Ile  
115 120 125  
Thr Phe Pro Asp Leu Trp Thr Asn Thr Cys Cys Ser His Pro Leu Cys  
130 135 140  
Ile Asp Asp Glu Leu Gly Leu Lys Gly Lys Leu Asp Asp Lys Ile Lys  
145 150 155 160  
Gly Ala Ile Thr Ala Ala Val Arg Lys Leu Asp His Glu Leu Gly Ile

165                    170                    175  
Pro Glu Asp Glu Thr Lys Thr Arg Gly Lys Phe His Phe Leu Asn Arg  
180                    185                    190  
Ile His Tyr Met Ala Pro Ser Asn Glu Pro Trp Gly Glu His Glu Ile  
195                    200                    205  
Asp Tyr Ile Leu Phe Tyr Lys Ile Asn Ala Lys Glu Asn Leu Thr Val  
210                    215                    220  
Asn Pro Asn Val Asn Glu Val Arg Asp Phe Lys Trp Val Ser Pro Asn  
225                    230                    235                    240  
Asp Leu Lys Thr Met Phe Ala Asp Pro Ser Tyr Lys Phe Thr Pro Trp  
245                    250                    255  
Phe Lys Ile Ile Cys Glu Asn Tyr Leu Phe Asn Trp Trp Glu Gln Leu  
260                    265                    270  
Asp Asp Leu Ser Glu Val Glu Asn Asp Arg Gln Ile His Arg Met Leu  
275                    280                    285

\*\*\*

SEQUENCE ID No.: 4

LENGTH: 1099

SEQUENCE TYPE: nucleic acid

STRANDNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: cDNA

ORIGIN

ORGANISM: Phaffia rhodozyma

STRAIN: ATCC 24230

SEQUENCE CHARACTERISTIC

CHARACTERISTIC CODE: CDS

LOCATION: 99..851

DETERMINATION METHOD: E

SEQUENCE:

CCCACGCGTC CGCACATCTC GCATATATCA CTTTCCTCCT TCCAGAACAA GTTCTGAGTC 60



185                    190                    195  
TTT GAG GAC GAG TCT AAC TCA TTT ACC CCT TGG TTC AAG TTG ATT GCC 740  
Phe Glu Asp Glu Ser Asn Ser Phe Thr Pro Trp Phe Lys Leu Ile Ala  
200                    205                    210  
CGA GAC TTC CTG TTT GGC TGG TGG GAT CAA CTT CTC GCC AGA CGA AAT 788  
Arg Asp Phe Leu Phe Gly Trp Trp Asp Gln Leu Leu Ala Arg Arg Asn  
215                    220                    225                    230  
GAA AAG GGT GAG GTC GAT GCC AAA TCG TTG GAG GAT CTC TCG GAC AAC 836  
Glu Lys Gly Glu Val Asp Ala Lys Ser Leu Glu Asp Leu Ser Asp Asn  
235                    240                    245  
AAA GTC TGG AAG ATG TAGTCGACC CTTCTTCTG TACAGTCATC TCAGTCGCC 890  
Lys Val Trp Lys Met \*\*\*  
250  
TGTTGGTTGC TTGCTTCTTG CTCTTCTTC TATATATCTT TTTCTTGCC TGGTAGACT 950  
TGATCTTCT ACATAGCATA CGCATACATA CATAAACTCT ATTTCTTGT TTTATCTCT 1010  
CTTCTAAGGG AATCTTCAAG ATCAATTCTT TTTGGGCTA CAACATTC GATCAATGTT 1070  
GCTTTTCAGA CTACAAAAAA AAAAAAAA 1099

SEQUENCE ID No.: 5

LENGTH: 1074

SEQUENCE TYPE: nucleic acid

STRANDNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: cDNA

ORIGIN

ORGANISM: Haematococcus pluvialis

STRAIN: NIES-144

SEQUENCE CHARACTERISTIC

CHARACTERISTIC CODE: CDS

LOCATION: 145..921

DETERMINATION METHOD: E

SEQUENCE:

ATCGCTACTT GGAACCTGGC CCGGCAGCAG TCCGATGACG CGATGCTTCG TTGTTGCTC 60  
AGAGGCCTCA CGCATTTCCC CCGCGTGAAC TCCGCGCAGC AGCCCAGCTG TGACACACGCG 120  
CGACTCCAGT TTAGGCCAG AAGC ATG CAG CTG CTT GCC GAG GAC CGC ACA GAC 179  
Met Gln Leu Leu Ala Glu Asp Arg Thr Asp 5 10  
CAT ATG AGG GGT GCA AGT ACC TGG GCA GGC GGG CAG TCG CAG GAT GAG 222  
His Met Arg Gly Ala Ser Thr Trp Ala Gly Gln Ser Gln Asp Glu  
15 20 25  
CTG ATG CTG AAG GAC GAG TGC ATC TTG GTG GAT GCT GAC GAC AAC ATT 270  
Leu Met Leu Lys Asp Glu Cys Ile Leu Val Asp Ala Asp Asp Asn Ile  
30 35 40  
ACA GGC CAT GTC AGC AAG CTG GAG TGC CAC AAG TTC CTA CCA CAT CAG 318  
Thr Gly His Val Ser Lys Leu Glu Cys His Lys Phe Leu Pro His Gln  
45 50 55  
CCT GCA GGC CTG CTG CAC CGG GCC TTC TCT GTA TTC CTG TTT GAC GAC 366  
Pro Ala Gly Leu Leu His Arg Ala Phe Ser Val Phe Leu Phe Asp Asp  
60 65 70  
CAG GGG CGA CTG CTG CAA CAG CGT GCA CGA TCA AAA ATC ACA TTC 414  
Gln Gly Arg Leu Leu Gln Gln Arg Ala Arg Ser Lys Ile Thr Phe  
75 80 85 90  
CCC AGT GTG TGG ACC AAC ACC TGC TGC AGC CAC CCT CTA CAT GGG CAG 462  
Pro Ser Val Trp Thr Asn Thr Cys Cys Ser His Pro Leu His Gly Gln  
95 100 105  
ACC CCA GAT GAG GTG GAC CAA CTA AGC CAG GTG GCC GAC GGC ACA GTA 510  
Thr Pro Asp Glu Val Asp Gln Leu Ser Gln Val Ala Asp Gly Thr Val  
110 115 120  
CCT GGC GCA AAG GCT GCT GCC ATC CGC AAG TTG GAG CAC GAG CTG GGG 558  
Pro Gly Ala Lys Ala Ala Ile Arg Lys Leu Glu His Glu Leu Gly  
125 130 135  
ATA CCA GCG CAC CAG CTG CCG GCC AGC GCG TTT CGC TTC CTC ACG CGT 606  
Ile Pro Ala His Gln Leu Pro Ala Ser Ala Phe Arg Phe Leu Thr Arg  
140 145 150

DRAFT - NOT FOR RELEASE

TTG CAC TAC TGC GCC GCG GAC GTG CAG CCG GCT GCG ACA CAA TCA GCA 654  
Leu His Tyr Cys Ala Ala Asp Val Gln Pro Ala Ala Thr Gln Ser Ala  
155 160 165 170  
CTC TGG GGC GAG CAC GAA ATG GAC TAC ATC TTA TTC ATC CGG GCC AAC 702  
Leu Trp Gly Glu His Glu Met Asp Tyr Ile Leu Phe Ile Arg Ala Asn  
175 180 185  
GTC ACC CTT GCG CCC AAC CCT GAC GAG GTG GAC GAA GTC AGG TAC GTG 750  
Val Thr Leu Ala Pro Asn Pro Asp Glu Val Asp Glu Val Arg Tyr Val  
190 195 200  
ACG CAG GAG GAG CTG CGG CAG ATG ATG CAG CCG GAC AAT GGG TTG CAA 798  
Thr Gln Glu Glu Leu Arg Gln Met Met Gln Pro Asp Asn Gly Leu Gln  
205 210 215  
TGG TCG CCG TGG TTT CGC ATC ATC GCC GCG CGC TTC CTT GAG CGC TGG 846  
Trp Ser Pro Trp Phe Arg Ile Ile Ala Ala Arg Phe Leu Glu Arg Trp  
220 225 230  
TGG GCT GAC CTA GAC GCG GCC CTG AAC ACT GAC AAA CAC GAG GAT TGG 894  
Trp Ala Asp Leu Asp Ala Ala Leu Asn Thr Asp Lys His Glu Asp Trp  
235 240 245 250  
GGA ACG GTG CAT CAC ATC AAC GAA GCG TGA AAACAG AAGCTGTAGG 940  
Gly Thr Val His Ile Asn Glu Ala \*\*\*  
255  
ATGTCAAGAC ACGTCATGAG GGGGCTTGGC ATCTTGGCGG CTTCGTATCT CTTTTACTG 1000  
AGACTGAACC TGCAGCTGGA GACAATGGTG AGCCCAATTG AACTTTCCGC TGCACTGGAA 1060  
AAAAAAAAAAA AAAA 1074

SEQUENCE ID No.: 6

LENGTH: 1058

SEQUENCE TYPE: nucleic acid

STRANNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: genomic DNA

ORIGIN

**ORGANISM:** *Saccharomyces cerevisiae*

**STRAIN: S288C**

## **SEQUENCE CHARACTERISTIC**

**CHARACTERISTIC CODE: CDS**

LOCATION: 187..1050

## DETERMINATION METHOD: S

## **SEQUENCE:**

ATA ACT TTC CCT GAT CTT TGG ACT AAC ACA TGC TGC TCT CAT CCA CTA 615  
 Ile Thr Phe Pro Asp Leu Trp Thr Asn Thr Cys Cys Ser His Pro Leu  
     130                        135                        140  
 TGT ATT GAT GAC GAA TTA GGT TTG AAG GGT AAG CTA GAC GAT AAG ATT 663  
 Cys Ile Asp Asp Glu Leu Gly Leu Lys Gly Lys Leu Asp Asp Lys Ile  
     145                        150                        155  
 AAG GGC GCT ATT ACT GCG GCG GTG AGA AAA CTA GAT CAT GAA TTA GGT 711  
 Lys Gly Ala Ile Thr Ala Ala Val Arg Lys Leu Asp His Glu Leu Gly  
     160                        165                        170                        175  
 ATT CCA GAA GAT GAA ACT AAG ACA AGG GGT AAG TTT CAC TTT TTA AAC 759  
 Ile Pro Glu Asp Glu Thr Lys Thr Arg Gly Lys Phe His Phe Leu Asn  
     180                        185                        190  
 AGA ATC CAT TAC ATG GCA CCA AGC AAT GAA CCA TGG GGT GAA CAT GAA 807  
 Arg Ile His Tyr Met Ala Pro Ser Asn Glu Pro Trp Gly Glu His Glu  
     195                        200                        205  
 ATT GAT TAC ATC CTA TTT TAT AAG ATC AAC GCT AAA GAA AAC TTG ACT 855  
 Ile Asp Tyr Ile Leu Phe Tyr Lys Ile Asn Ala Lys Glu Asn Leu Thr  
     210                        215                        220  
 GTC AAC CCA AAC GTC AAT GAA GTT AGA GAC TTC AAA TGG GTT TCA CCA 903  
 Val Asn Pro Asn Val Asn Glu Val Arg Asp Phe Lys Trp Val Ser Pro  
     225                        230                        235  
 AAT GAT TTG AAA ACT ATG TTT GCT GAC CCA AGT TAC AAG TTT ACG CCT 951  
 Asn Asp Leu Lys Thr Met Phe Ala Asp Pro Ser Tyr Lys Phe Thr Pro  
     240                        245                        250                        255  
 TGG TTT AAG ATT ATT TGC GAG AAT TAC TTA TTC AAC TGG TGG GAG CAA 999  
 Trp Phe Lys Ile Ile Cys Glu Asn Tyr Leu Phe Asn Trp Trp Glu Gln  
     260                        265                        270  
 TA GAT GAC CTT TCT GAA GTG GAA AAT GAC AGG CAA ATT CAT AGA ATG 1047  
 Leu Asp Asp Leu Ser Glu Val Glu Asn Asp Arg Gln Ile His Arg Met  
     275                        280                        285  
 CTA TAA CAACG 1058  
 Leu \*\*\*